This article describes an interesting application of Markov chain Monte Carlo (MCMC). MCMC is used to assess competing explanations of marital fertility decline. Data collected during the World Fertility Study in Iran are analyzed using methods developed to perform discrete time event history analyses in which unobserved heterogeneity is explicitly accounted for. The usual age-period-cohort identifiability problem is compounded by the presence of a fourth clock, duration since previous birth, and a fifth clocklike variable, mother's parity. The authors resolve this problem by modeling some of the clocks parametrically using codings suggested by alternating conditional expectation (ACE) and Bayes factors to decide which clocks are necessary. Compound Laplace-Metropolis estimates are used to compute Bayes factors for comparing alternative models. The new methods enable the authors to conclude that Iran's fertility decline was primarily a period effect and not a cohort effect, that it started before the Family Planning Program was initiated, that it was the same for women at all educational levels but varied depending on husband's education, and that it was greatest in the largest cities, particularly Tehran.

Bayesian Analysis of Event History Models With Unobserved Heterogeneity via Markov Chain Monte Carlo

Application to the Explanation of Fertility Decline

STEVEN M. LEWIS
ADRIAN E. RAFTERY
University of Washington

1. INTRODUCTION

With the world's population now greater than 5.5 billion, the need for reducing human fertility is as great as ever. Although fertility rates in most developed countries have dropped dramatically during the

AUTHORS' NOTE: *This research was supported by National Institutes of Health Grant 5-R01-HD26330. We are grateful to Akbar Aghajanian for providing expert consultation throughout our study of human fertility in Iran; Dr. Jamshidi, director of the Iran Statistical Center, for giving us permission to analyze the Iran Fertility Survey data; Jamshidi and Dr. Nejatian for encouragement and support; and two anonymous reviewers for helpful comments.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 28 No. 1, August 1999 35-60

past century (Coale and Watkins 1986), high fertility rates are still common in many developing countries. The demographic question of what triggers and drives a fertility decline remains an open one.

Even before World War II, economic modernization was proposed by many as the best predictor of reduced human fertility. In formulating the classical theory of the demographic transition, Notestein et al. (1944) and other leading demographers of their time argued that economic modernization, in the form of increased industrialization and urbanization, would lead to a substantial decrease in human fertility. As elaborations on the theory of the demographic transition, Becker (1960) proposed his demand theory and Easterlin and Crimmins (1985) presented a theoretical framework combining both the demand for and the supply of children. In the context of any of these three theoretical frameworks, economic modernization remained central to reducing the growth rate of the human population.

Based on the findings of the European Fertility Project (Coale and Watkins 1986), alternative *ideational* explanations for reduced fertility rates received attention in the demographic literature (e.g., Cleland and Wilson 1987). Instead of emphasizing economic modernization, ideation theory focused more on the diffusion of ideas as an explanation for fertility decline. It was the increased knowledge and acceptance of fertility control that was most important in reducing fertility. Shared social customs, culture, and language were proposed as important variables in explaining reduced fertility.

There is still an ongoing debate on which of these theoretical explanations for reduced human fertility is more useful. This article describes how we addressed this question using event history models in which unobserved heterogeneity is explicitly accounted for.

A large amount of data on human fertility and many potential explanatory variables have been accumulated since World War II. This includes the World Fertility Survey (WFS) (for an overview, see Cleland and Scott 1987); numerous knowledge, attitude, and practice (KAP) surveys (Bongaarts 1991; International Institute for Population Studies 1972); and the Demographic and Health Surveys (DHS) (IRD/Macro International 1991). Iran was a participant in the WFS, but because of the 1979 revolution, the data from Iran were not included in most of the results published from the WFS.

To use event history models in our study of what leads to decreased human fertility, we found it necessary to use a number of new and recently developed statistical methods; methods that were not available at the time when the WFS was completed. These new methods include the following:

- use of discrete time event history analysis with unobserved heterogeneity
- use of event history analysis to evaluate inclusion of alternative demographic clock variables and to determine whether parity needs to be included
- use of substantively meaningful variables to represent the period effect
- use of the alternating conditional expectation (ACE) technique to suggest how to code some of the covariates
- interpretation of the coefficients found using event history analysis in terms of the total fertility rate
- use of Bayes factors to compare alternative models and approximation of the Bayes factors using the Compound Laplace-Metropolis estimator.

The use of these methods in the application of event history analysis to the Iranian data is the primary focus of this article.

2. FINDINGS

In our study of the Iran Fertility Survey (IFS), the Iranian portion of the WFS, we found that fertility drops as women become more educated and, to a lesser extent, as the father becomes more educated. Another strong predictor of reduced fertility is whether previous children are still living at the time of the survey. A lesser predictor is the size of the place of current residence. Our results using the Iranian data corroborated the results found in previous studies of human fertility, such as Bumpass, Rindfuss, and Palmore (1986), Hobcraft (1985), Gilks (1986), and Cleland and Rodríguez (1988). However, using the methods discussed here, we were able to find a number of new results:

- The Iranian fertility decline was a period effect and not a cohort effect, equally affecting childbearing women of all ages.
- The period effect on the fertility decline could be parsimoniously coded using the level of primary school participation.

- Parity was associated with fertility at all parity levels, not just for first and second births, contrary to Hobcraft's (1985) findings.
- The Iranian fertility decline started before the Family Planning Program was initiated.
- The fertility decline was the same for women at all educational levels, but its extent depended on husband's education.
- The fertility decline was largest in Tehran, substantial but less than in Tehran in other urban areas, and relatively small in rural areas and small towns.
- The onset of the fertility decline was preceded by a fertility increase.

Detailed discussions of these findings may be found in Raftery, Lewis, and Aghajanian (1995).

Our models also included a number of demographic control variables, which we refer to as clocks. We explored models with up to five different types of clocks. These may be described as follows:

1. age of the mother
2. duration since previous birth or since initial union
3. mother's cohort
4. period (i.e., calendar year)
5. mother's parity.

Thus, the usual age-period-cohort identifiability problem was present and was compounded by the presence of a fourth clock, duration since previous birth, that appears in event history data but not in data with one observation per individual, and a fifth clocklike variable, mother's parity. We resolved the identifiability problem by modeling some of the clocks (age, duration, parity) parametrically using codings suggested by the ACE method (Breiman and Friedman 1985) when substantively plausible and supported by the data. As an example, we found that duration of birth interval was best coded using dummy variables for intervals of 0, 1, or 2 years and a continuous variable using an exponential transformation for intervals longer than 2 years. More details on the use of ACE and on how each of the clocks were coded may be found in Raftery, Lewis, Aghajanian, and Kahn (1996). We then used Bayes factors to determine which clocks were needed in the model.

We found that age, duration, and parity were significant in all of our models. However, we observed that including both cohort variables and period variables in the model was not as good as using the period variables alone. Hence, we concluded that the onset of the fertility decline in Iran was mostly a period effect and not a cohort effect. This finding is discussed further in Section 4.3.

We were also able to substantially reduce the number of period variables needed by replacing our original set of period variables by a variable using the proportion of eligible children who actually attended primary school. This variable was one of a number of variables that we investigated for representing the period effect. Other variables studied included the enactment of the Family Protection Act in 1967 and the rate of participation in the Family Planning Program, which was initiated in 1967. Neither of these variables represented the period effect nearly as well as the rate of primary participation. This finding has implications for the broader debate on the causes of fertility decline. See Section 4.3 for more information about the different alternatives we explored for coding the period effect for use in our event history models.

3. INITIAL ANALYSES

3.1. DATA

The WFS is a set of individual surveys carried out in 42 developing countries between 1972 and 1984. The data collected included full retrospective pregnancy histories, marital histories, contraceptive use data, data on breast-feeding of most recent children, and selected socioeconomic characteristics. Analyses of the data collected were completed for all but one of the developing countries. The one country was Iran. Analysis of the IFS, which was carried out in 1976 and 1977, was not completed because of the Iranian revolution of 1979.

The IFS was based on a nationally representative sample of ever-married women who were less than 50 years old at the time of the survey. Based on a multistage random sampling procedure, 6,056 households were visited and all ever-married women less than 50 years old were interviewed. This resulted in interviews being completed with I = 4,912 women. There were a total of 20,641 births in 93,006 ever-married-women years.

Aghajanian, Gross, and Lewis (1993) assessed the quality of the IFS data. They found that the data were of generally good quality, at least as good and in many ways better than data from other WFS participating countries. The types of errors found in the IFS were similar to errors found in other WFS surveys. There were a modest number of misreported ages and misreported dates of events such as births, deaths, and so on. The single biggest problem was that although respondents were asked for both the month and year when events occurred, the month frequently could not be provided. Therefore, we worked with only the year in which events took place. We did experiment with imputation methods for month of birth, but the results were almost the same. This probably reflects the fact that the month was missing in roughly two thirds of the cases.

3.2. EXPLORATORY ANALYSES

Before performing more elaborate analyses, we did a variety of initial explorations of the IFS data. These were done to find which variables were the most likely to be associated with fertility, as well as to look into how to best code the variables. These initial explorations consisted of linear regressions in which the response was the number of years between successive births, in effect treating each birth interval as an independent outcome even though most women had more than one interval; censored birth intervals were omitted. As an initial exploratory technique, these regressions proved to be very informative. Included as independent variables in these regressions were variables such as the mother's parity (number of previous children), mother's birth cohort, mother's age at marriage, mother's work status (however, in Iran most mothers do not work outside the home), size of the place of mother's current residence, both mother's and father's level of completed education, and so on.

Based on these regressions, we determined that including separate indicator variables for each parity level was unnecessary. We found that using a single indicator variable for parity 1 and one other variable linear in parity was adequate. We also noticed that the set of independent variables associated with parity 0 intervals was quite different than those associated with parity 1 and greater intervals. Because our purpose was to investigate variables associated with reducing fertility, it made more sense to develop a model for parity 1 and greater intervals and to exclude parity 0 intervals from further study. In the IFS data set, there were 16,997 parity 1 and greater births in 77,279 ever-married-women years.

Also, we found that using indicator variables for mother's birth cohort (grouped in 5-year intervals) improved the model. Mother's work status was found to be insignificant. Mother's age at marriage was associated with birth intervals only for first children and not with parity 1 and greater intervals, so there is no need to consider it further in this article. Both mother's and father's level of education contributed significantly to the model.

While running these regressions, we also used the ACE technique of Breiman and Friedman (1985) to check on possible monotonic transformations of independent variables. ACE suggested that a monotonic transformation of the size of place of current residence variable was called for. This enabled us to reduce the number of categories for the size variable from eight to five. An example of the use of this technique is given in Section 4.4.

4. METHOD

4.1. DISCRETE TIME EVENT HISTORY ANALYSIS USING LOGISTIC REGRESSION

We modeled the marital fertility histories in the IFS using discrete time event history analysis (Allison 1984; Yamaguchi 1991). To apply event history analysis to the IFS data, each year that a woman was married is treated as a separate case. We refer to each of these cases as a single woman exposure year.

The discrete time event history model may be estimated using logistic regression:

$$\begin{aligned} \operatorname{logit}(\pi_{it}) &= \log\left(\frac{\pi_{it}}{1-\pi_{it}}\right) \\ &= \beta_0 + \sum_{p=1}^{P} \beta_p x_{pit}, \end{aligned} \tag{1}$$

where π_{it} is the probability that the *i*th woman has a child in the *t*th year of a birth interval, x_{pit} is the *p*th covariate for the *i*th woman in year *t*, and $\beta_0, \beta_1, \ldots, \beta_P$ are unknown regression coefficients. Logistic regression can be used to estimate discrete time event history models by treating each discrete unit of exposure as if it were an independent observation (Allison 1984; Holford 1976).

We included control variables to take account of demographic differences between exposure years. That is, we included independent variables for demographic clocks such as current age, duration since previous birth, parity, birth cohort, and/or period (i.e., calendar year). Before any further development of a discrete event history model for fertility, we had to determine how best to code each of these five clocks. We also had to determine which of these clocks to include in the model.

We did this by comparing models with various subsets of the clock variables using Bayes factors, approximated by the Bayesian information criterion (BIC) statistic, which is defined as

$$BIC = -\chi^2 + P \log N, \tag{2}$$

where χ^2 is the likelihood ratio test statistic for comparing the null model with no covariates to the model of interest, *P* is the number of independent variables in the model of interest (not counting the intercept) as defined by equation (1), and *N* is the sample size—that is, the number of cases (exposure years) in the logistic regression (Kass and Raftery 1995; Raftery 1995). BIC is an approximation to twice the log Bayes factor for the null model against the model being considered; the approximation is particularly good for a specific reasonable choice of prior (Kass and Wasserman 1995). The smaller BIC is (i.e., the more negative), the better the model. For the model with no covariates, BIC is zero, so a positive BIC indicates a model that is worse than the null model. This criterion also has the intuitively appealing feature that it combines a measure of absolute goodness of fit (χ^2) with a penalty for the number of parameters ($P \log N$).

4.2. CODING AGE AND DURATION USING PARAMETRIC MODELS

There is a clear relationship between fertility and age of the mother. To assess the functional form of this, we calculated the number of

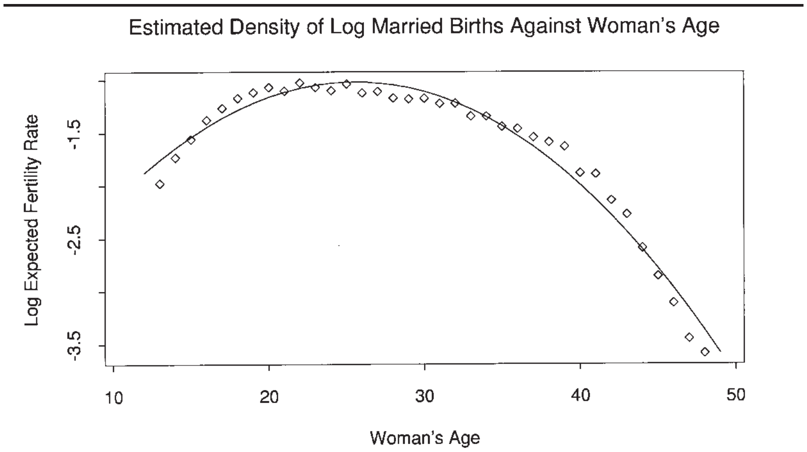

Figure 1: Estimated Log Fertility Rate With Fitted Quadratic Polynomial

woman exposure years and the number of births to women of that age in the IFS data set for each year of age below 50. Dividing the number of births by the number of woman exposure years gives an estimate of the average age-specific marital fertility rate. The logarithm of the fertility rate is fit well by a quadratic polynomial function of woman's age; namely,

$$\log(\mathit{fertility}) = -1.155 - 0.509a - 0.468a^2 \, (R^2 = 0.93), \tag{3}$$

where $a = (age - mean(age))/10$. In Figure 1, the estimated log fertility rate is shown together with the fitted curve given by equation (3). The curve fits well.

We were also able to find a good parametric coding of the duration since previous birth variable, which we denote by t. Figure 2 shows a plot of the empirical fertility rate in the Iranian data set, calculated using the method just described for age. Starting in the third year following a birth, the empirical fertility rate declines almost exactly exponentially (R^2 = .99 in a weighted logarithmic regression). Hence, it makes sense to use four covariates to code duration: one dummy variable each for durations t = 0, 1, and 2, and the remaining variable defined as

$$\begin{cases} 0 & \mathit{if} \quad t = 0,1,2 \\ \operatorname{logit}(0.726 \times 0.795^t) & \mathit{if} \quad t \geq 3 \end{cases} \tag{4}$$

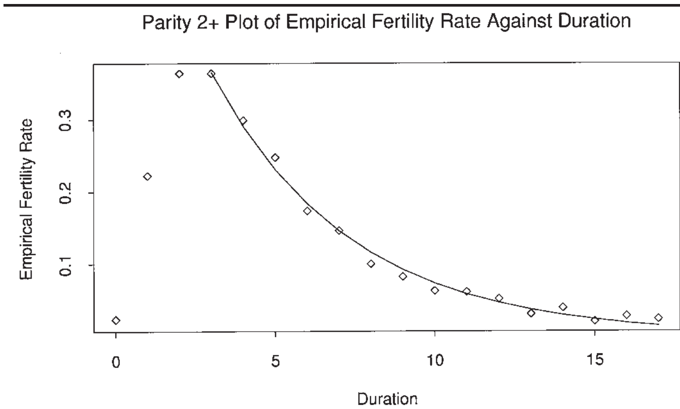

Figure 2: Estimated Fertility Rate by Duration

to represent the exponential portion of the density, where the two constants in the argument to the logit function were found empirically from the Iranian data set.

4.3. WAS THE FERTILITY DECLINE A PERIOD EFFECT, A COHORT EFFECT, OR BOTH?

Table 1 contains log likelihoods and BIC values for logistic regression models containing various subsets of the clock variables. It is clear that mother's age, duration since previous birth, and parity belong in the model.

Models 4 and above include birth cohort variables and period variables. Six indicator variables, one for each 5-year interval between ages 20 and 50, were used to code cohort; women younger than 20 were the baseline cohort. Similarly, six indicator variables, one for each 5-year interval between 1943 and 1972, were used to code period; exposure years between 1973 and 1977 were the baseline period. BIC becomes more negative when either the cohort variables or the period variables are added, so that one or the other should be included. However, BIC becomes less negative when both sets of variables are included, showing that cohort and period should *not* both be kept in the model. Because BIC for model 5 is less than BIC for model 4, we concluded that keeping period in the model is preferred over keeping cohort in the model. In other words, when the period effect is in the model, the cohort effect accounts for very little

TABLE 1: Logistic Regressions Using Different Demographic Clock Variables

Model	*Clock Variables*	χ^2	P	*BIC*
0	Null	0	0	0
1	A	1,243	2	–1,220
2	A D	12,587	6	–12,519
3	A D B	12,678	8	–12,588
4	A D B C	12,761	14	–12,603
5	*A D B Y*	*12,839*	*14*	*–12,682*
6	A D B C Y	12,850	20	–12,625

NOTE: The preferred model is shown in italics. The independent variables are as follows: A = age effect (equation (3)), D = duration since previous birth (equation (4)), B = parity (number of previous births) and parity 1 indicator, C = cohort (seven levels), Y = period (seven levels). χ^2, *P*, and BIC are defined by equation (2).

additional variability (for additional discussion of this finding, see Raftery, Lewis, and Aghajanian 1995).

In Table 1, we used six variables to code the period effect. In Raftery, Lewis, and Aghajanian (1995), we detailed how we were able to find a more parsimonious and substantively meaningful way to code the period effect. We found that the period effect was well represented by a piecewise linear function with a change point in 1959. We coded the period effect using just two variables. The first was coded as the calendar year for those exposure years before 1959 and as constant for years after 1959. The second was coded as constant for years before 1959 and as the percentage primary school participation for years after 1959. We considered several other variables besides primary school participation for representing the post-1959 period effect. Some of the variables were justified by the classical theory of the demographic transition, whereas others were justified by appealing to ideational explanations of reduced fertility rates. Among the variables justified by the theory of the demographic transition were gross domestic product (GDP), primary school participation (PRIM) and secondary school enrollments (SEC). We considered two variables associated with ideational changes that took place in Iran. The official Family Planning Program (FPP) started in 1967, and the Iranian Family Protection Act (FPA) became law that same year. We found that of the three demographic transition variables, PRIM explained the period effect best. Neither GDP nor SEC fit the period effect as well as a simple linear predictor (i.e., calendar year). Neither of the ideational variables

(FPP, FPA) did much better than a model consisting of only an intercept, and a simple linear predictor was much better than either variable.

4.4. CODING OF INDEPENDENT VARIABLES USING ACE

We had no reason to believe that all of the independent variables considered for use in our initial linear regressions contributed to the model in a linear fashion. Perhaps the effects of one or more of the variables were nonlinear. We used the ACE technique of Breiman and Friedman (1985) to check for nonlinear effects of independent variables in the IFS data.

A number of the independent variables were categorical on either an ordered or interval scale; these included the educational attainment variables and the size of place of current residence variable. Variables such as these are often found to contribute in a nonlinear fashion in linear regression models. In this section, we show how ACE was used to find an improved coding, taking as an example the size of place of current residence variable.

Size of place of current residence was originally coded in eight categories ranging from Tehran (1) to isolated farm dwellings (8). The ACE transformation is shown in Figure 3. This is fairly linear between categories 1 and 5, but there is almost no change between categories 5 and 8. Categories 6 through 8 are all rural, whereas category 5 consists of small towns. In fact, most of the small towns in category 5 are more rural than urban and are actually large villages, with more than 5,000 inhabitants. We amalgamated categories 5 to 8 to form a single category. This indicates that there are no systematic differences in marital fertility between rural and small-town residents in terms of the size of their place of residence, but for city dwellers the size of the place where they live does have an effect. This resulted in coding the size of place of current residence linearly as a variable taking on values 1 through 5.

4.5. ACCOUNTING FOR UNOBSERVED HETEROGENEITY USING MCMC

In previous studies, several researchers have observed that event history analysis results can be misleading unless any unobserved heterogeneity is accounted for in the estimation (Heckman and Singer

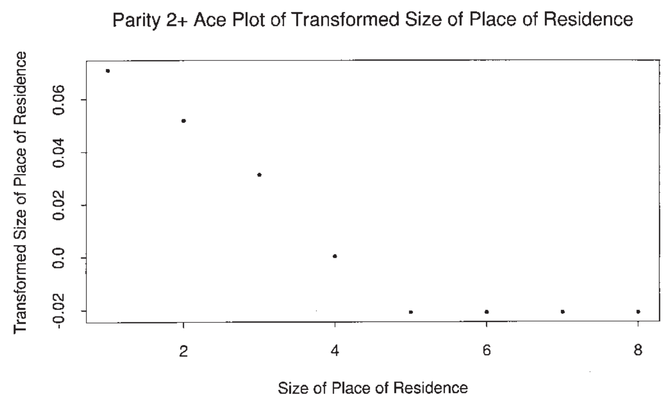

Figure 3: ACE Plot for Size of Place of Current Residence

1984). It has been argued by Heckman and Singer and other researchers that it is safer to assume the presence of unobserved heterogeneity unless it is shown that there is none. To do so requires adding parameters to the logistic regression model (equation (1)) to account for any unobserved differences between units of observation. In the IFS, the units of observation were the individual women. So, to check for any unobserved heterogeneity, we fitted models of the form

$$\operatorname{logit}(\pi_{it}) = \beta_0 + \sum_{p=1}^{P} \beta_p x_{pit} + \alpha_i, \tag{5}$$

where α_i were unobserved woman-specific random effects representing unmeasured characteristics that affect fertility, such as fecundability and coital frequency. We assumed that α_i were independent random variates from a normal distribution with a mean of zero and a common variance, σ^2. Adding α_i captures sources of unobserved heterogeneity that may induce correlation across spells for the same person.

Models of the form of equation (5) can no longer be estimated using standard logistic regression. In addition to the fixed effects, $\beta_0, \ldots, \beta_P$, we also needed to estimate all α_i and σ^2. These can be estimated simultaneously using a Bayesian estimation procedure, such as Markov chain Monte Carlo (MCMC) (Hastings 1970; Smith and Roberts 1993). To use MCMC for this purpose, we had to find an appropriate set of univariate posterior conditional distributions for the parameters, at least up to an unknown normalizing constant.

We assumed that the prior distribution of the fixed effects was a diffuse, but proper, multivariate-normal distribution, with its mass well spread out over those parts of the parameter space that were at all probable (see Raftery 1996). We assumed that the prior distribution of the random effects was also normal. In the usual mixed-effects model (i.e., one containing both fixed and random effects), it is common practice to assume that the variance of the random effects, σ^2, has an inverted gamma prior distribution. We followed this practice. These prior distributions are detailed in Appendix A, which also provides a rationale for the values of the prior hyperparameters used.

It is straightforward to combine the logistic likelihood with the prior distributions to arrive at univariate posterior conditional distributions for each of the parameters. These are shown in Appendix B. These actual posterior conditionals, not normal approximations, were then used in MCMC to estimate all of the parameters simultaneously.

5. RESULTS

5.1. ADDING INDIVIDUAL CHARACTERISTICS TO THE MODEL

So far, we have looked only at models containing different subsets of the demographic clock variables. But to address our original goal of comparing alternative explanations for changes in human fertility, we had to consider various variables reflecting these alternatives. In other words, we needed to compare models in which variables such as mother's and father's education, child mortality, use of contraception, and so on were added to the model. In Table 2, we compare several of these models.

Models including use of contraception were not as good as the models shown in Table 2. Although it is conceivable that use of contraception is not predictive of reduced fertility, other explanations are available. The contraception variables collected during the IFS asked only whether the women were knowledgeable about various contraceptive methods and whether they had *ever* used different methods. Just knowing about contraception may not have much of an

TABLE 2: Logistic Regressions Using Different Individual Characteristic Variables

Model	*Other Variables*	χ^2	P	*BIC*
0	Only clocks	12,893	11	–12,769
1	E	13,447	12	–13,311
2	E F	13,532	13	–13,386
3	E F S	13,647	14	–13,490
4	*E F S M*	*14,544*	*15*	*–14,375*

NOTE: All models include age, duration, parity, and period. The preferred model is shown in italics. The independent variables are as follows: E = mother's completed education (five categories), F = father's completed education (five categories), S = size of the place in which the woman resides (five categories), M = child mortality (1 if the previous child was alive, 0 if not). χ^2, *P*, and BIC are defined by equation (2).

association with a drop in fertility. Also, whether a woman ever used contraception is probably not precise enough to be predictive. It would have been helpful to know how regularly the women used contraception and whether the requisite contraceptive guidelines were adhered to, but this information was not available.

Up to this point, no allowance for any unobserved heterogeneity had been made. We estimated the model that allows for the presence of any unobserved heterogeneity (equation (5)) by running MCMC for 18,500 iterations. The last 18,000 iterations were retained, discarding the first 500 iterations for burn-in. These MCMC control parameters were selected using the *gibbsit* method and software developed by Raftery and Lewis (1992, 1996). This consists of computing the number of iterations needed to achieve a given level of precision in estimating specific posterior quantiles, such as the .025 and .975 points that define a 95% Bayesian confidence interval for a regression coefficient. This is done by reducing the chain to a two-state (binary) sequence and applying standard Markov chain theory. Note that this method provides answers to questions of when the algorithm has converged to the region of parameter space favored by the posterior distribution and how long it should be run *after* it has reached the right region. These are quite different questions, although they are often confused with one another. We also tried a number of different starting values but found that the results were not noticeably affected by these.

The model consisted of the same fixed effects as model 4 in Table 2 and a random effect parameter for each woman in the sample. Table 3 shows the posterior means of the fixed-effect parameters (found from

MCMC), their posterior standard deviations, the associated *t* values (i.e., the posterior mean divided by its posterior standard deviation), and 95% highest posterior density intervals. The null deviance was 81,427 on 77,278 degrees of freedom. The deviance for the model was 65,253, so $\chi^2 = 16{,}174$. All of the variables retained in the model were highly significant. Both of the education variables, the size of the place of current residence and child mortality, were found to be important predictors of human fertility behavior.

What was the effect of including heterogeneity in the model? For the model with fixed effects alone, the deviance was 66,883 on 77,263 degrees of freedom, so χ^2 was 14,544 on 15 degrees of freedom. The increase in χ^2 when heterogeneity was added was 1,630. There were 4,912 women in the sample, so the effect of unobserved heterogeneity in our analysis of the IFS data was not large. The posterior means in Table 3 are close to fixed-effect parameter estimates found using logistic regression alone (not shown), which is consistent with our finding of minimal effect of unobserved heterogeneity in the IFS data. Thus, our conclusions were relatively insensitive to any unobserved heterogeneity. This is not guaranteed always to be true, however. Indeed, it is generally agreed among demographers that unobserved heterogeneity *is* present in data such as these, due to factors such as biological differences and differences in marital coital frequency. Thus, it is wise to account for unobserved heterogeneity explicitly when modeling such data.

There is another benefit to using MCMC to get a sample from the posterior distribution. The sample can be used to examine marginal posterior distributions of the fixed effects graphically, permitting the researcher to check for any asymmetry or nonnormality of any of the marginal distributions. This can be done using a nonparametric density estimation technique such as the one described by Terrell (1990). We did so for all of the fixed effects in Table 3 and found that the marginal posterior distributions were symmetric and approximately normally distributed.

We then considered models with interaction variables. We found the best model (based on BIC) to be a model with 10 interaction variables in addition to the variables in Table 3. All of the duration variables had significant interactions with mother's education. Duration 0 interacted with parity. The exponential transformation of durations

TABLE 3: Estimates for the Preferred Model from Table 2

				95% Interval	
Variable	*Estimate*	SE	t *Value*	*Lower*	*Upper*
Intercept	–2.70	0.20	–13.5	–2.99	–2.33
Age (linear)	–0.37	0.03	–14.3	–0.42	–0.32
Age (quadratic)	–0.31	0.02	–17.9	–0.34	–0.28
Duration 0	–2.39	0.07	–33.3	–2.53	–2.25
Duration 1	0.34	0.05	6.4	0.24	0.45
Duration 2	1.22	0.05	23.1	1.12	1.32
Duration 3+	2.42	0.10	24.2	2.23	2.62
Parity 1	0.23	0.03	7.4	0.17	0.29
Parity	–0.08	0.01	–10.3	–0.10	–0.07
Pre-1959 linear trend	0.04	0.003	11.3	0.03	0.04
Post-1959 primary participation	–0.40	0.08	–4.9	–0.56	–0.24
1977	–0.25	0.05	–4.7	–0.35	–0.14
Woman's education	–0.18	0.02	–10.2	–0.22	–0.15
Husband's education	–0.06	0.01	–4.9	–0.09	–0.04
Size of place of residence	–0.07	0.01	–8.6	–0.08	–0.05
Child mortality	–0.76	0.03	–29.6	–0.81	–0.71

longer than 2 years (see Section 4.2) interacted with post-1959 primary participation. Post-1959 primary participation also interacted with the size of place of current residence and with father's education. Parity 1 interacted with mother's education. Also, there was a significant interaction between the size of place of current residence and father's education.

5.2. MODEL COMPARISONS USING BAYES FACTORS

The standard method for comparing models within the Bayesian framework is to calculate Bayes factors (for a survey, see Kass and Raftery 1995). The Bayes factor, B_{01}, for comparing model M_0 with model M_1 for observed data, $\mathbf{Y}$, is the ratio of the posterior odds for M_0 against M_1 to the prior odds, which reduces to

$$B_{01} = \frac{f(\mathbf{Y}|\,M_0)}{f(\mathbf{Y}|\,M_1)}.$$

In other words, the Bayes factor is the ratio of the integrated (or marginal) likelihoods of the two models being compared. Hence, calculation of Bayes factors boils down to computing integrated likelihoods,

$$f(\mathbf{Y}|M_m) = \int f(\mathbf{Y}|\theta_m, M_m) f(\theta_m|M_m) d\theta_m \qquad m = 0,1,$$

where θ_m is the vector of parameters in model M_m and $f(\theta_m \mid M_m)$ is its prior density. Dropping the notational dependence on the model, this can be rewritten as

$$f(\mathbf{Y}) = \int f(\mathbf{Y}|\theta) f(\theta) d\theta. \tag{6}$$

Calculating exact Bayes factors for comparing hierarchical models is difficult in general. Instead, it is necessary to estimate integrated likelihoods under each of the alternative models. The output from MCMC can be used to do this. Lewis and Raftery (1997) described how MCMC output can be used to estimate log-integrated likelihoods for logistic hierarchical models, such as those presented in this article, using the Compound Laplace-Metropolis estimator. For logistic hierarchical models, the estimator may be written as

$$\widehat{\mathcal{LM}}_c = \frac{P+1}{2}\log\{2\pi\} + \frac{1}{2}\log\{|\mathbf{H}^*|\} + \log\{f(\theta^*)\} + \sum_{i=1}^{I} \hat{\mathcal{L}}_i, \tag{7}$$

where θ^* is the value of θ at which $h \equiv \log\{f(\mathbf{Y} \mid \theta)f(\theta)\}$ attains its maximum, $\mathbf{H}^*$ is minus the inverse Hessian of h evaluated at θ^*, and

$$\begin{aligned}\hat{L}_i = &-\frac{1}{2}\log\left(1+\tilde{\sigma}^2\left[\sum_t \frac{\exp\{\mathbf{X}_{it}\tilde{\eta}+\alpha_i^*\}}{(1+\exp\{\mathbf{X}_{it}\tilde{\eta}+\alpha_i^*\})^2}\right]\right) - \left(\frac{1}{2\tilde{\sigma}^2}\right)\alpha_i^{*2} \\ &+\left[\sum_t y_{it}(\mathbf{X}_{it}\tilde{\eta}+\alpha_i^*)\right] - \left[\sum_t \log(1+\exp\{\mathbf{X}_{it}\tilde{\eta}+\alpha_i^*\})\right].\end{aligned}$$

$\mathbf{X}_{it}$ is the vector of covariates for the ith woman at her tth exposure year; $\tilde{\eta}$ and $\tilde{\sigma}^2$ are the estimated fixed-effect parameters and variance of the random effect parameters, respectively, at the joint mode of the posterior distribution; and α_i^* is the mode of

TABLE 4: Using Compound Laplace-Metropolis to Compute Bayes Factors

Model	*Other Variables*	$\hat{\mathcal{LM}}_c$	SE ($\hat{\mathcal{LM}}_c$)	*log*(B_{0m})
0	Only clocks	–34,275	1.24	0
4	*E F S M*	*–33,510*	*2.49*	*–765*
5	E F S M FPP	–33,518	2.34	–757

NOTE: All models include age, duration, parity, and period. The preferred model is shown in italics. The independent variables are as follows: E = mother's completed education (five categories), F = father's completed education (five categories), S = size of the place in which the woman resides (five categories), M = child mortality (1 if the previous child was alive, 0 if not), FPP = official family planning program (1 if yes, 0 if no). $\hat{\mathcal{LM}}_c$ is defined by equation (7).

$$h_2(\alpha_i) \equiv \left\{ \left[\sum_t y_{it}(\mathbf{X}_{it}\tilde{\eta} + \alpha_i) \right] - \left(\frac{1}{2\tilde{\sigma}^2} \right) \alpha_i^2 - \left[\sum_t \log(1 + \exp\{\mathbf{X}_{it}\tilde{\eta} + \alpha_i\}) \right] \right\}$$

for the *i*th woman with random effect parameter α_i. For logistic hierarchical models, this Compound Laplace-Metropolis estimator provides very good estimates of the integrated likelihoods.

In Table 4, we show Compound Laplace-Metropolis estimates of integrated likelihoods under models 0 and 4 from Table 2 and for a model with FPP added to model 4. The standard errors of the Compound Laplace-Metropolis estimates shown in Table 4 were found using the method of batch means (Lewis and Raftery 1997). By taking differences between Compound Laplace-Metropolis estimates, we computed log Bayes factors for comparing the models in Table 4. The choice of which model to use as the baseline is arbitrary. In Table 4, model 0 was used as the baseline. The log Bayes factor for comparing model 4 to model 5 is readily determined from the numbers in Table 4. It is (–765 – (–757)) = –8, showing that model 4 fits better than either of the other two models. This happens to be the same model as was preferred in Table 2, but here we have accounted for any unobserved heterogeneity that might have been present. In particular, it is noteworthy that the FPP variable added in model 5 led to a model that did not fit as well as model 4.

6. DISCUSSION

In this article, we have presented a successful application of MCMC in a practical application where it had not been previously used. We found that MCMC worked well for estimating discrete time event history models with unobserved heterogeneity included. Although we found only a small amount of unobserved heterogeneity in the Iran data, it is reassuring to know that it was accounted for and that the significance of the main effects was real. This permits us to confidently discuss the conclusions these main effects imply.

We found that the death of a previous child was the most significant individual characteristic leading to an increased probability of having another child (Table 3). The next most significant individual characteristic was how educated the woman was. Her husband's level of education was also important, but the *t* value associated with husband's education was only half that associated with the woman's education. The other significant individual characteristic we found was how large the city or town where the woman currently resided was. These results provide additional support to prior research findings such as Coale and Watkins (1986) and D'Souza (1974). In particular, D'Souza's most significant finding was the importance of the death of a previous child. Also in D'Souza's book is an extensive discussion of a model for fertility rate by duration assuming a combined normal-exponential distribution for fertility, as compared to the empirically derived method we used in Section 4.2.

During our investigation, we explored many models including other possible covariates. We fitted models using variables such as knowledge and use of various forms of contraception, duration of breast-feeding, or kind of work the woman was employed in (if she was employed), as well as a number of different family-planning variables. Inclusion of one or more of these variables did not improve the model, based on the Bayes factor comparing the model presented in Section 4.5 with any larger model.

APPENDIX A
Prior Distributions

We assumed that the prior distribution of the fixed-effect parameters, β, was multivariate normal with mean vector **m** and covariance matrix **V**, where $\mathbf{V} = \mathbf{QUQ^T}$, with

$$\mathbf{Q} = (s_0) \begin{bmatrix} 1 & -\bar{x}_2 / s_2 & -\bar{x}_3 / s_3 & \cdots & -\bar{x}_P / s_P \\ 0 & 1/s_2 & 0 & \cdots & 0 \\ 0 & 0 & 1/s_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/s_P \end{bmatrix},$$

where $\bar{x}_p = \sum_{i,t} w_{it} x_{pit} / \sum_{i,t} w_{it}$ is the weighted sample mean of the pth covariate, $s_p^2 = \sum_{i,t} w_{it} (x_{pit} - \bar{x}_p)^2 / \sum_{i,t} w_{it}$ is the weighted sample variance of the pth covariate, $s_0^2 = \sum_{i,t} w_{it} (z_{it} - \bar{z})^2 / \sum_{i,t} w_{it}$ is the weighted sample variance of the adjusted dependent variable, $z_{it} = \text{logit}\,(\hat{\pi}_{it}) + (y_{it} - \hat{\pi}_{it}) \frac{d}{d\pi} \text{logit}\,(\hat{\pi}_{it})$, with weights $w_{it} = \hat{\pi}_{it}(1 - \hat{\pi}_{it})$, $\mathbf{m} = (\bar{z}, 0, \ldots, 0)^T$ and $\mathbf{U} = \text{Diag}\{\psi^2, \phi^2, \ldots, \phi^2\}$. This form of prior was suggested by Raftery (1996), who proposed setting ψ^2 to 1 and ϕ^2 to 1.65. We followed this suggestion here.

In other words, we assumed that the joint prior distribution for the vector of fixed-effect parameters, β, was

$$f(\beta) \propto \left[\exp\left\{ -\frac{1}{2} (\beta - \mathbf{m})^T \mathbf{V}^{-1} (\beta - \mathbf{m}) \right\} \right].$$

The random effect parameters, α_i, were assumed to be independent normals with a common variance parameter, σ^2. So, the prior distribution of a single random effect parameter would be simply

$$f(\alpha_i | \sigma^2) \propto \left[\exp\left\{ -\left(\frac{1}{2\sigma^2} \right) \alpha_i^2 \right\} \right].$$

The prior distribution for the variance of the random effects, σ^2, was assumed to be an inverted gamma distribution. It is clearer expressing this distribution in terms of the precision parameter, $\tau^2 = 1/\sigma^2$. The prior may then be written as

$$f(\tau^2) \propto (\tau^2)^{(r-1)} \exp\{-\zeta \tau^2\} I_{(0,\infty)}(\tau^2),$$

where r and ζ are hyperparameters. Thus, the precision τ^2 has an ordinary gamma distribution.

What values should be used for the two hyperparameters? Lewis (1994) found that the empirical 99% quantile of the number of children ever born to women who were at least 45 years old in the IFS was 13; the same result was found if all women who were at least 40 years old were included. Raftery, Lewis, and Aghajanian (1995) found the empirical mean birth rate in the IFS was about 0.22 births per woman exposure year. Hence, over an average length reproductive lifetime of about 30 years, the average Iranian woman would be expected to have about $30 \times 0.22 \approx 6.6$ children. So, at least approximately, we would expect $\text{logit}\left(\frac{13}{30}\right) - \text{logit}(0.22) = 0.9974$ to be about Φ^{-1} (0.99) = 2.3263 standard units from the mean, where Φ is the cumulative distribution function of a standard normal random variable. In other words, standard deviations of random effects greater than $\left(\frac{0.9974}{2.3263}\right) = 0.429$ units on the logit scale, or equivalently a variance of about 0.184, would be rather unlikely based on what we know about human fertility behavior. So, a reasonable upper limit for the variance was $U \equiv 0.184$; that is, roughly 1/5.

Obviously, zero was a lower limit for the variance. However, if the variance of the random effects was only negligibly greater than zero, it would not be of any substantive interest, and so we did not want the prior to assign too much probability to these negligible variances. We wanted to be able to detect random deviations of even as little as 3/10 child per woman. In other words, we wanted to be able to detect differences as small as

$$\text{logit}\left(\frac{6.6}{30}\right) - \text{logit}\left(\frac{6.6 - 0.3}{30}\right) = \text{logit}\ (0.22) - \text{logit}(0.21) = 0.05926 \approx 1/16,$$

which is equivalent to a variance of only 1/256. Because we wanted the prior distribution of the variance of the random effects to not exclude the possibility of a variance as small as 1/256, a reasonable lower limit was $L \equiv 1/256$. So we wanted a prior that assigned nearly all of its probability to values of σ^2 between 1/256 and 0.184.

Between these two limits, we wanted the prior to be as flat as we could make it. We operationalized this by bounding the ratio between the maximum and minimum ordinates of an inverted gamma density over the interval between the two limits. Following arguments made by Jeffreys (1961), we bounded this ratio by $B \equiv 10$. We used this upper bound for the ratio to derive two nonlinear inequality constraints for the two hyperparameters. Appropriate values for the two hyperparameters were found by locating the largest r and ζ satisfying these constraints. Lewis (1994) provides details on how this was done. For the case in which $L = 1/256$, $U = 0.184$, and $B = 10$, we found $r = 0.47$ and $\zeta = 0.0226$. Figure 4 shows what an inverted gamma distribution with these shape and scale parameters looks like.

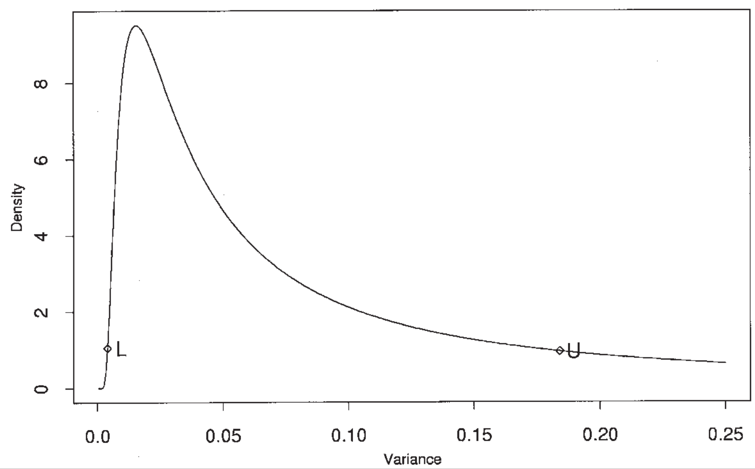

Figure 4: Inverted Gamma Density With Shape Parameter, *r* = 0.47, and Scale Parameter, ζ = 0.0226

APPENDIX B
Univariate Posterior Conditional Distributions

At least up to the normalizing constant, each of the univariate posterior conditional distributions may be found by combining the logistic likelihood and the appropriate prior (for derivations of the following results, see Lewis 1994).

The joint posterior conditional distribution of the fixed-effect parameters is

$$f(\beta|\alpha,\mathbf{Y}) \propto \left[\prod_{i,t} \frac{\left(e^{\mathbf{X}_{it}\beta+\alpha_i}\right)^{y_{it}}}{1+e^{\mathbf{X}_{it}\beta+\alpha_i}}\right] \exp\left\{-\frac{1}{2}\sum_{r=1}^{P}\sum_{s=1}^{P}(\beta_r - m_r)v_{r,s}(\beta_s - m_s)\right\}, \tag{8}$$

where $\mathbf{Y}$ is a vector of the observed outcomes, m_r is the rth element of the mean vector $\mathbf{m}$, and $v_{r,s}$ is the entry in the rth row and the sth column of the *inverse* of the hypercovariance matrix, $\mathbf{V}$.

During the MCMC run, we needed the univariate posterior conditional distribution for a single fixed-effect parameter. This may be derived from equation (8) after a fair amount of algebra. For the pth fixed effect, the univariate posterior conditional distribution is

$$f(\beta_p|\beta_{-p},\alpha,\mathbf{Y})$$
$$\propto \left[\prod_{i,t}\frac{\left(e^{\mathbf{X}_{-pit}\beta_{-p}+\alpha_i}\right)^{yit}}{1+e^{\mathbf{X}_{-pit}\beta_{-p}+\alpha_i}}\right]$$
$$\exp\left\{-(\beta_p-m_p)\left[\frac{1}{2}v_{p,p}(\beta_p-m_p)+\sum_{s\in\mathcal{P}}v_{p,s}(\beta_s-m_s)\right]\right\},$$

where β_{-p} is the vector of fixed-effect parameters excluding the fixed effect whose conditional distribution is being calculated; $\mathbf{X}_{-pit}$ is the itth row of the covariate matrix $\mathbf{X}$, with the pth column excluded; and $\mathcal{P}=1,\ldots,p-1,p+1,\ldots,P$.

The univariate posterior conditional distribution for a random effect parameter is

$$f(\alpha_i|\beta,\tau^2,\mathbf{Y}_i)\propto\left[\prod_t\frac{\exp\{\mathbf{X}_{it}\beta+\alpha_i\}^{y_{it}}}{1+\exp\{X_{it}\beta+\alpha_i\}}\right]\left[\left(\frac{\tau^2}{2\pi}\right)^{\frac{1}{2}}\exp\left\{-\left(\frac{\tau^2}{2}\right)\alpha_i^2\right\}\right],$$

and the univariate posterior conditional distribution for the precision of the random effects is

$$f(\tau^2|\alpha)\propto(\tau^2)^{(\frac{I}{2}+r-1)}\exp\left\{-\left(\zeta+\frac{1}{2}\sum_{i=1}^{I}\alpha_i^2\right)\tau^2\right\}I_{(0,\infty)}(\tau^2).$$

REFERENCES

Aghajanian, A., A. B. Gross, and S. M. Lewis. 1993. "Evaluation of Iran Fertility Survey." Working Paper No. 93-2, University of Washington, Center for Studies in Demography and Ecology.

Allison, P. D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. Sage University Paper Series on Quantitative Applications in the Social Sciences No. 07-046. Beverly Hills, CA: Sage.

Becker, G. S. 1960. "An Economic Analysis of Fertility." Pp. 209-31 in *Demographic and Economic Change in Developed Countries*. Princeton, NJ: Princeton University Press.

Bongaarts, J. 1991. "The KAP-Gap and the Unmet Need for Contraception." *Population and Development Review* 17:293-313.

Breiman, L., and J. H. Friedman. 1985. "Estimating Optimal Transformations for Multiple Regression and Correlation (With Discussion)." *Journal of the American Statistical Association* 8:580-619.

Bumpass, L. L., R. R. Rindfuss, and J. A. Palmore. 1986. "Determinants of Korean Birth Intervals: The Confrontation of Theory and Data." *Population Studies* 40:403-23.

Cleland, J. G., and G. Rodríguez. 1988. “The Effect of Parental Education on Marital Fertility in Developing Countries.” *Population Studies* 42:419-42.

Cleland, J. G., and C. Scott, eds. 1987. *The World Fertility Survey: An Assessment*. Oxford: Oxford University Press.

Cleland, J. G., and C. Wilson. 1987. “Demand Theories of the Fertility Transition: A Iconoclastic View.” *Population Studies* 41:5-30.

Coale, A. J., and S. C. Watkins, eds. 1986. *The Decline of Fertility in Europe: The Revised Proceedings of a Conference on the Princeton European Fertility Project*. Princeton, NJ: Princeton University Press.

D’Souza, S. 1974. *Closed Birth Intervals: A Data Analytic Study.* New Delhi: Sterling.

Easterlin, R. A., and E. M. Crimmins. 1985. *The Fertility Revolution: A Supply-Demand Analysis*. Chicago: University of Chicago Press.

Gilks, W. R. 1986. “The Relationship Between Birth History and Current Fertility in Developing Countries.” *Population Studies* 40:437-55.

Hastings, W. K. 1970. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications.” *Biometrika* 57:97-109.

Heckman, J. J., and B. Singer. 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica* 52:271-320.

Hobcraft, J. N. 1985. “Family-Building Patterns.” Pp. 64-86 in *Reproductive Change in Developing Countries*, edited by J. G. Cleland and J. Hobcraft. London: Oxford University Press.

Holford, T. R. 1976. “Life Tables With Concomitant Information.” *Biometrics* 32:587-97.

International Institute for Population Studies. 1972. *Knowledge, Attitude and Practice (KAP) Surveys in Family Planning: An Overview of Two Decades of Research*. Bombay: International Institute for Population Studies.

IRD/Macro International. 1991. *Proceedings of the Demographic and Health Surveys World Conference*. 3 vols. Washington, DC: IRD/Macro International.

Jeffreys, H. 1961. *Theory of Probability.* 3d ed. Oxford: Oxford University Press.

Kass, R. E., and A. E. Raftery. 1995. “Bayes Factors.” *Journal of the American Statistical Association* 90:773-95.

Kass, R. E., and L. Wasserman. 1995. “A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion.” *Journal of the American Statistical Association* 90:928-34.

Lewis, S. M. 1994. “Multilevel Modeling of Discrete Event History Data Using Markov Chain Monte Carlo Methods.” Ph.D. thesis, University of Washington, Department of Statistics.

Lewis, S. M., and A. E. Raftery. 1997. “Estimating Bayes Factors via Posterior Simulation With the Laplace-Metropolis Estimator.” *Journal of the American Statistical Association* 92:648-55.

Notestein, F. W., I. B. Taeuber, D. Kirk, A. J. Coale, and L. K. Kiser. 1944. *The Future Population of Europe and the Soviet Union: Population Projections 1940-1970*. Geneva: League of Nations.

Raftery, A. E. 1995. “Bayesian Model Selection in Social Research (With Discussion).” Pp. 111-96 in *Sociological Methodology 1995*, vol. 25, edited by P. V. Marsden. Cambridge, MA: Blackwell.

———. 1996. “Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models.” *Biometrika* 83:251-66.

Raftery, A. E., and S. M. Lewis. 1992. “How Many Iterations in the Gibbs Sampler?” Pp. 763-73 in *Bayesian Statistics 4*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A.F.M. Smith. Oxford: Oxford University Press.

———. 1996. “Implementing MCMC.” Pp. 115-30 in *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. London: Chapman & Hall.

Raftery, A. E., S. M. Lewis, and A. Aghajanian. 1995. “Demand or Ideation? Evidence From the Iranian Marital Fertility Decline.” *Demography* 32:159-82.

Raftery, A. E., S. M. Lewis, A. Aghajanian, and M. J. Kahn. 1996. “Event History Modeling of World Fertility Survey Data.” *Mathematical Population Studies* 6:129-53.

Smith, A.F.M., and G. O. Roberts. 1993. “Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods (With Discussion).” *Journal of the Royal Statistical Society, Series B* 55:3-102.

Terrell, G. R. 1990. “The Maximal Smoothing Principle in Density Estimation.” *Journal of the American Statistical Association* 85:470-77.

Yamaguchi, K. 1991. *Event History Analysis*. Applied Social Research Methods Series Vol. 28. Newbury Park, CA: Sage.

Steven M. Lewis is a research associate in the School of Social Work at the University of Washington. He has published articles on statistics, demography, and applied social research in the leading journals of those fields.

Adrian E. Raftery is a professor of statistics and sociology at the University of Washington. He is currently working on Bayesian model selection and Bayesian model averaging in social research (www.research.att.com/~volinsky/bma.html), model-based clustering (www.stat.washington.edu/fraley/resources.html), inference for mechanistic models in whale population dynamics and environmental monitoring, and family structure and social mobility.